

AVERAGING AND DETERMINISTIC OPTIMAL CONTROL*

F. CHAPLAIS†

Abstract. Averaging is often used in ordinary differential equations when dealing with fast periodic phenomena. It is shown here that it can be used efficiently in optimal control. As the period tends to zero, a limit or “averaged” problem is defined. The open loop optimal control of the limit problem induces a cost which is optimal up to the second order when evaluated through the original dynamics. The definition of the averaged problem is then generalized to the nonperiodic case. It is shown that the Bellman function of the original “fast” problem tends uniformly on any compact set to that of the averaged problem.

Key words. averaging, optimal control, time scales, perturbations in control, Hamilton–Jacobi equations

AMS(MOS) subject classifications. 34C29, 41A60, 4900, 49C20

Introduction: A perturbation approach. Averaging can be seen as part of the perturbation theory of differential equations. Consider

$$(1) \quad \frac{dx}{dt} = f^\varepsilon(x, t), \quad x(0) = x_0, \quad x(t) \in \mathbf{R}^n, \quad t \in [0, T].$$

Regular perturbations correspond to the situation when f^ε has a limit in $C^0(\mathbf{R}^n \times [0, T], \mathbf{R}^n)$ as ε tends to zero. Singular perturbations [6] can be seen as f^ε having a limit in $L^2(\mathbf{R}^n \times [0, T], \mathbf{R}^n)$. Roughly speaking, averaging is the case when f^ε has a limit in $L^2(\mathbf{R}^n \times [0, T], \mathbf{R}^n)$ in the weak topology. As an example, consider

$$(2) \quad \frac{dx}{dt} = f\left(x, t, \frac{t}{\varepsilon}\right)$$

where f is periodic in the last variable. Define $f^\varepsilon(x, t) = f(x, t, t/\varepsilon)$; clearly f^ε has no limit either pointwise or in L^2 ; yet f^ε tends to f^0 defined by $f^0(x, t) = 1/\omega \int_0^\omega f(x, t, \theta) d\theta$, weakly in L^2 .

It is well known [1] that the solution of (2) can be approximated by the solution of

$$(3) \quad \frac{dy}{dt} = f^0(y, t), \quad y(0) = x_0$$

with an error of order $\varepsilon\omega$, provided that f be regular enough. Solving (3) instead of (2) is known as “averaging”.

What is good for differential equations is often good for optimal control. Regular and singular perturbations have been extensively studied ([2], [6]). As far as averaging is concerned, it has been studied in the context of partial differential equations [3] or stochastic optimal control [5]. We present here some approximation results in deterministic optimal control.

1. The periodic case.

1.1. The averaged problem. Let

$$f: \mathbf{R}^n \times \mathbf{R}^p \times [0, T] \times \mathbf{R} \rightarrow \mathbf{R}^n, \\ (x, u, t, \theta) \rightarrow f(x, u, t, \theta),$$

* Received by the editors February 3, 1986; accepted for publication (in revised form) May 5, 1986.

† Centre d'Automatique et Informatique de l'Ecole Nationale Supérieure des Mines de Paris, 35, rue Saint-Honoré, 77305 Fontainebleau Cedex, France.

$$L: \mathbf{R}^n \times \mathbf{R}^p \times [0, T] \times \mathbf{R} \rightarrow \mathbf{R},$$

$$(x, u, t, \theta) \rightarrow L(x, u, t, \theta),$$

with f and L periodic in θ with a period ω independent of x, u and t (regularity will be considered in § 1.2). Let U^{ad} be a constraint domain and $W^{ad} = \{u \in L^2([0, T], \mathbf{R}^p), u(t) \in U^{ad} \text{ for almost every } t\}$ the set of admissible controls. We define the problem (P^ε) as:

$$(4) \quad \left. \begin{aligned} \frac{dx}{dt} &= f\left(x(t), u(t), t, \frac{t}{\varepsilon}\right), \quad x(0) = x_0, \\ \text{Minimize } \int_0^T L\left(x(t), u(t), t, \frac{t}{\varepsilon}\right) dt &\text{ in } W^{ad}, \end{aligned} \right\} (P^\varepsilon).$$

We define the associated problem \bar{P} as follows:

$$(5) \quad \left. \begin{aligned} V^{ad} &= \{v \in L^2([0, T] \times [0, \omega], \mathbf{R}^p), v(t, \theta) \in U^{ad} \text{ a.e. } (t, \theta)\}, \\ \frac{dy}{dt} &= \frac{1}{\omega} \int_0^\omega f(y(t), v(t, \theta), t, \theta) d\theta, \quad y(0) = x_0, \\ \text{Minimize } \int_0^T \frac{dt}{\omega} \int_0^\omega L(y(t), v(t, \theta), t, \theta) d\theta &\text{ in } V^{ad}, \end{aligned} \right\} (\bar{P}).$$

Problem (4) can be seen as the perturbation of (5) by a fast oscillating input of null average. Notice that if f is Lipschitz in x and measurable in u, t, θ , then, for v in V^{ad} , the average \bar{f} of f as defined in (5) is Lipschitz in x and measurable in t . Hence, both (4) and (5) have a unique solution over $[0, T]$.

Remark 1. At first sight, it would seem reasonable to consider in (5) the averages of f and L in θ independently of u , that is, u being a constant vector in \mathbf{R}^p and not a function in $L^2([0, \omega], \mathbf{R}^p)$. This amounts to restricting V^{ad} to controls which are constant over $[0, \omega]$ at time t . As we shall see, this may lead to a severe loss of optimality when ε tends to 0. In short, it is necessary to have a feedback on the fast time θ . Consider for instance the following problem: $n = p = 1$, $U^{ad} = \mathbf{R}$, $f(x, u, t, \theta) = -x + u \sin(\theta)$ and $L(x, u, t, \theta) = x^2 + (u^2/6)$. If u is independent of θ , the average of f is equal to $-x$, which is itself independent of u . Within this class of functions, the optimal cost for (P^ε) is asymptotically equal to $\bar{J}(0)$, if $\bar{J}(u)$ denotes the cost of u in (\bar{P}) . $\bar{J}(0)$ is equal to $x_0^2(1 - e^{-2T})/2$.

Now use our definition of the averaged problem to compute a better control. Let q be the solution of the Riccati equation of the averaged problem: $dq/dt = 3q^2 + 2q - 1$, $q(T) = 0$; then $q(0) = (1 - e^{-4T})/(3 + e^{-4T})$. Define z by $dz/dt = -z(1 + 3q)$, $z(0) = x(0)$; z is the optimal trajectory for the averaged problem. Now use the open-loop control $u^\varepsilon(t) = -6q(t)z(t) \sin(t/\varepsilon)$ in problem (P^ε) . Then it is easy to see that x^ε driven by u^ε in (4) is asymptotic to z . Hence the cost of u^ε in (P^ε) is asymptotic to $\int_0^T z^2(t) \times (1 + 3q^2(t)) dt$. Thanks to the Riccati equation, the latter quantity is equal to $q(0)x^2$, which is smaller than $\bar{J}(0)$; thus for ε small enough, u^ε is better than any "slow" control.

Notice that the same phenomenon can be observed in stochastic optimal control, where it is well known that open-loop controls alone are not enough to ensure optimality. In both cases this is due to the presence of averages w.r.t. the events or fast time in the problem (the same can be said of the "ordinary," "slow" time, of course).

Remark 2. Even though the controls of problem (\bar{P}) are in an infinite dimensional state, the minimum principle and dynamic programming apply; the important fact is that the state is finite-dimensional. One also checks that the Bellman function is the viscosity solution of the Hamilton-Jacobi-Bellman equation [7].

Let $\bar{f}(x, v, t)$ denote the average of f and \bar{L} the average of L . The adjoint state equations for the problem (\bar{P}) are:

$$\frac{dp}{dt} = -\frac{\partial \bar{f}^T}{\partial x} p - \frac{\partial \bar{L}}{\partial x}, \quad p(T) = 0$$

and if v^* is an optimal control for the problem (\bar{P}) , the minimum principle says that $v^*(t, \cdot)$ minimizes the Hamiltonian of (\bar{P}) , that is, $p(t)^T \bar{f}(x, v, t) + \bar{L}(x, v, t)$, with respect to v in V^{ad} . This is equivalent to: $v^*(t, \theta)$ minimizes $p(t)^T f(x, v, t, \theta) + L(x, v, t, \theta)$ with respect to v in U^{ad} , this for almost every θ . We see that the control appears naturally as a feedback on the fast time (cf. Remark 1).

Remark 3. (\bar{P}) is better conditioned numerically than (P^ε) , since (P^ε) involves a time grid of order $\delta t/\varepsilon$, while (\bar{P}) involves only a grid of order δt , at least in the simulation part. A time grid in $\delta t/\varepsilon$ is still needed to compute the averages and to minimize the Hamiltonian.

1.2. An approximation theorem. The averaged problem is used here to compute a near optimal control for the problem (P^ε) , with an error of order ε^2 . The proof and assumptions are close to those used in [2].

Assumptions.

- (H1) $U^{ad} = \mathbf{R}^p$; no constraint.
- (H2) L does not depend on θ ; f and L are C^0 , of class C^2 in x , and u ; the first- and second-order derivatives of f are bounded, and Lipschitz; the second-order derivatives of L are bounded, and Lipschitz.
- (H3) (\bar{P}) has a solution, with optimal control u_0 , trajectory y and adjoint state q .
- (H4) Let $H(p, x, u, t, \theta) = p^T f(x, u, t, \theta) + L(x, u, t, \theta)$. There exists $\beta > 0$ such that, for all (x, u, t, θ) , $\partial^2 H / \partial v^2$ is greater than β Identity at point $(q(t), x, u, t, \theta)$, in the sense of the cone of positive semidefinite matrices.
- (H5) One has

$$\left[\frac{\partial^2 H}{\partial x^2} - \frac{\partial^2 H}{\partial x \partial v} \left(\frac{\partial^2 H}{\partial v^2} \right)^{-1} \frac{\partial^2 H}{\partial v \partial x} \right] \geq 0 \quad \text{at } (q(t), x, u, t, \theta) \text{ for all } (x, u, t, \theta).$$
- (H6) $f(y(t), u_0(t, \theta), t, \theta)$ and $\frac{\partial H}{\partial x}(q(t), y(t), u_0(t, \theta), t, \theta)$ are C^1 in t with a Lipschitz derivative.

Assumptions (H4) and (H5) ensure sufficient regularity of the control with respect to the cost. Assumption (H6) is specific to averaging: if it did not hold, there would not be a true separation of time scales.

THEOREM 1. *Let f and L meet Assumptions (H1)–(H6). Define u^ε by $u^\varepsilon(t) = u_0(t, t/\varepsilon)$. Then u^ε is near optimal for the problem (P^ε) with an error on the cost of order ε^2 .*

Sketch of proof. (A detailed proof can be found in [4].) We will proceed in two stages. First, we will exhibit a lower bound of the cost for any control u . We will then show that the control u^ε induces a cost which approximates this lower bound with an error of order ε^2 . It is then easy to conclude that u^ε is near optimal for the problem (P^ε) .

Remark 4. Unless otherwise stated, all partial derivatives will be taken at time t , with x being $y(t)$, u being $u_0(t, t/\varepsilon) = u^\varepsilon(t)$, θ being t/ε and the adjoint state being $q(t)$. We will make use of the following conventions: if $g(\sigma, \theta)$ is a periodic function

of θ , we will denote by $\bar{g}(\sigma)$ or $\bar{g}(\sigma, \cdot)$ or $\text{av}(g(\sigma, \cdot))$ the average of g in θ . We then define the operator Π on periodic functions g , by $\Pi(g(\sigma, \cdot))$ being the only primitive in θ of $g - \bar{g}$ with a null average. Π plays a key role in all developments of integrals or solutions of differential equations involving a periodicity in fast time. Finally, the superscript T denotes transposition.

LEMMA 1. Let $x_2(t, \theta) = \Pi(f(y(t), u_0(t, \cdot), t, \cdot)(\theta))$. There exists $\varepsilon_0 > 0$, $k \geq 0$, such that, for any control u and any $\varepsilon \in]0, \varepsilon_0[$, x being the trajectory driven by u in (4), the following estimation holds:

$$\begin{aligned} \int_0^T L(x(t), u(t), t) dt &\geq \int_0^T L\left(y(t), u_0\left(t, \frac{t}{\varepsilon}\right), t\right) dt \\ &\quad + \varepsilon \int_0^T \text{av} \left[\frac{\partial H}{\partial x}(y(t), q(t), u_0(t, \cdot), t, \cdot) x_2(t, \cdot) \right] dt \\ &\quad - \varepsilon q(0)^T x_2(0, 0) - k\varepsilon^2. \end{aligned}$$

Proof of Lemma 1. We will denote by \tilde{x} and \tilde{u} the following errors:

$$\tilde{x}(t) = x(t) - y(t) - \varepsilon x_2\left(t, \frac{t}{\varepsilon}\right), \quad \tilde{u}(t) = u(t) - u_0\left(t, \frac{t}{\varepsilon}\right).$$

It should be noticed that, if x_1 is defined by

$$(6) \quad \frac{dx_1}{dt} = \frac{\partial \tilde{f}}{\partial x} x_1 + \frac{\partial \tilde{f}}{\partial x} x_2, \quad x_1(0) + x_2(0, 0) = 0$$

and if $\tilde{u} = 0$ (that is, x is the trajectory driven by u^ε), then $y + \varepsilon x_1 + \varepsilon x_2(t, t/\varepsilon)$ is a uniform approximation of x with an ε^2 error.

We will use developments of functions with integral remains. To this end, we will use the following notation: for λ in $[0, 1]$, μ in $[0, 1]$, $F_t(\lambda, \mu)$ will denote the Hessian symmetric operator associated with the second-order derivatives of f in x and u , at point $(y(t), \lambda \mu(\varepsilon x_2(t, t/\varepsilon) + \tilde{x}(t)), u_0(t, t/\varepsilon) + \lambda \mu \tilde{u}(t), t, t/\varepsilon)$. $L_t(\lambda, \mu)$ will denote the analogue for L and $H_t(\lambda, \mu) = q^T(t) F_t(\lambda, \mu) + L_t(\lambda, \mu)$ will denote the Hessian of H at the same point.

We will also make use of the following linear quadratic oscillatory “tangent” problem:

$$(7) \quad \left. \begin{aligned} \frac{dz}{dt} &= \frac{\partial f}{\partial x} \left(z + x_2\left(t, \frac{t}{\varepsilon}\right) \right) + \frac{\partial f}{\partial u} v, \quad z(0) + x_2(0, 0) = 0, \\ \text{Min}_v \int_0^T &\left[\frac{1}{2} (z^T, v^T) H_t(0, 0) (z^T, v^T)^T + p_2^T \left(t, \frac{t}{\varepsilon} \right) \left(\frac{\partial f}{\partial x} z + \frac{\partial f}{\partial u} v \right) \right] dt, \end{aligned} \right\} (TP^\varepsilon)$$

where $p_2 = -\Pi(\partial H^T / \partial x)$. p_2 is the analogue of x_2 for the adjoint state.

Thanks to (H4) and (H5), (TP^ε) has a unique solution with trajectory y_1 , optimal control v_1 and adjoint state q_1 . Moreover, $\|y_1\|_\infty$ and $\|v_1\|_\infty$ are bounded when ε ranges within a neighbourhood of zero. Finally estimates will bear on the quantity

$$z_\varepsilon^2 = \int_0^1 \lambda d\lambda \int_0^1 d\mu \|Z_\varepsilon(\lambda, \mu)\|_{L^2}^2$$

where

$$\begin{aligned} Z_\varepsilon(\lambda, \mu)(t) &= v(t) + \left[\left(\frac{\partial^2 H}{\partial v^2} v \right)^{-1} \frac{\partial^2 H}{\partial v \partial x} \right] \left(r(t) + \varepsilon x_2\left(t, \frac{t}{\varepsilon}\right) \right), \\ r &= \tilde{x} - \varepsilon y_1, \quad v = \tilde{u} - \varepsilon v_1, \end{aligned}$$

the derivatives being taken at the same interpolation points as for $F_t(\lambda, \mu)$, $L_t(\lambda, \mu)$ and $H_t(\lambda, \mu)$.

At last, \approx will denote an approximation with an ε^2 error.

LEMMA 1.1.

$$\begin{aligned} \int_0^T L(x, u, t) dt &\approx \int_0^T L(y, u^\varepsilon, t) dt + \varepsilon \int_0^T \overline{\frac{\partial H}{\partial x}} x_2 dt - \varepsilon q(0)^T x_2(0, 0) \\ &\quad + \int_0^T \left[\frac{\partial H}{\partial x} - \frac{\partial \bar{H}}{\partial x} \right] \tilde{x} dt \\ &\quad + \int_0^T dt \int_0^1 \lambda d\lambda \int_0^1 d\mu \left(\tilde{x}^T + \varepsilon x_2^T \left(t, \frac{t}{\varepsilon} \right), \tilde{u}^T \right) H_t(\lambda, \mu) \\ &\quad \times \begin{pmatrix} \tilde{x} + \varepsilon x_2(t, t/\varepsilon) \\ \tilde{u} \end{pmatrix}. \end{aligned}$$

Proof of Lemma 1.1. The cost is expanded at the second order using an integral remain. Since there is no constraint, $q^T(\partial f/\partial u) + (\partial L/\partial u) = 0$; the Hamiltonian appears after a classical integration by parts. We then neglect all integrals depending on fast time when of order larger than or equal to ε^2 .

LEMMA 1.2. *There exists $k \geq 0$ such that*

$$\|r\|_\infty^2 \leq k(\varepsilon^2 + z_\varepsilon^2) \quad \text{and} \quad \|v\|_{L^2}^2 \leq k(\varepsilon^2 + z_\varepsilon^2).$$

Proof of Lemma 1.2.

$$\begin{aligned} \frac{dr}{dt} &= \frac{dx}{dt} - \frac{dy}{dt} - \varepsilon \frac{d}{dt} \left(x_2 \left(t, \frac{t}{\varepsilon} \right) \right) - \varepsilon \frac{dy_1}{dt} \\ (8) \quad &= f \left(r + y + \varepsilon x_2 + \varepsilon y_1, v + u_0 + \varepsilon v_1, t, \frac{t}{\varepsilon} \right) - f \left(y, u_0 \left(t, \frac{t}{\varepsilon} \right), t, \frac{t}{\varepsilon} \right) - \varepsilon \frac{\partial f}{\partial x} (y_1 + x_2) \\ &\quad - \varepsilon \frac{\partial f}{\partial u} v_1 - \varepsilon \Pi \left[\frac{\partial}{\partial t} (f(y, u_0(t, \cdot), t, \cdot)) \right] \left(\frac{t}{\varepsilon} \right), \end{aligned}$$

since $\partial/\partial\theta[\Pi(f)] = f - \bar{f}$. The second expression is equal to

$$\begin{aligned} &f \left(r + y + \varepsilon x_2 + \varepsilon y_1, v + u_0 + \varepsilon v_1, t, \frac{t}{\varepsilon} \right) - f \left(y + \varepsilon x_2 + \varepsilon y_1, u_0 + \varepsilon v_1, t, \frac{t}{\varepsilon} \right) \\ &\quad + \varepsilon^2 \int_0^1 \lambda d\lambda \int_0^1 d\mu (x_2^T + y_1^T, v_1^T) F_t(\lambda, \mu) \begin{pmatrix} x_2 + y_1 \\ v_1 \end{pmatrix} - \varepsilon \Pi \left[\frac{\partial}{\partial t} (f) \right] \left(\frac{t}{\varepsilon} \right). \end{aligned}$$

v then appears as a difference in the controls and is replaced by

$$Z_\varepsilon(\lambda, \mu) - \left[\left(\frac{\partial^2 H}{\partial v^2} \right)^{-1} \frac{\partial^2 H}{\partial v \partial x} \right] (r + \varepsilon x_2).$$

Using the Gronwall lemma, and neglecting the integrals of fast periodic functions at the second order, we get the estimate on $\|r\|_\infty$. The estimation on $\|v\|_{L^2}$ is obtained through the definition of Z_ε .

LEMMA 1.3. *r can be approximated uniformly with an error of order ε^2 by r_1 , with*

$$\begin{aligned} (9) \quad \frac{dr_1}{dt} &= f \left(r_1 + y + \varepsilon x_2 + \varepsilon y_1, v + u_0 + \varepsilon v_1, t, \frac{t}{\varepsilon} \right) - f \left(y + \varepsilon x_2 + \varepsilon y_1, u_0 + \varepsilon v_1, t, \frac{t}{\varepsilon} \right), \\ r_1(0) &= 0. \end{aligned}$$

In particular, one has $\|r_1\|_\infty^2 \leq 2k(\varepsilon^2 + z_\varepsilon^2)$.

Proof of Lemma 1.3. Note that (9) is close to (8). We then get the result by using the Gronwall lemma. We will now use r_1 rather than r .

LEMMA 1.4. *There exists $k \geq 0$ such that, for $\varepsilon \in]0, 1]$:*

$$\int_0^T \left(\frac{\partial H}{\partial x} - \frac{\partial \bar{H}}{\partial x} \right) \tilde{x} dt \geq \varepsilon \int_0^T p_2^T \left(t, \frac{t}{\varepsilon} \right) \left[\frac{\partial f}{\partial x} r_1 + \frac{\partial f}{\partial u} v \right] dt - k\varepsilon^2 - k\varepsilon z_\varepsilon^2.$$

Proof of Lemma 1.4.

$$-\varepsilon \frac{d}{dt} \left(p_2^T \left(t, \frac{t}{\varepsilon} \right) \right) = \varepsilon \Pi \left[\frac{\partial}{\partial t} \left(\frac{\partial H}{\partial x} (q(t), y(t), u_0(t, \cdot), t, \cdot) \right) \right] \left(\frac{t}{\varepsilon} \right) + \frac{\partial H}{\partial x} - \frac{\partial \bar{H}}{\partial x}$$

and $\tilde{x} \approx r_1 + \varepsilon y_1$. Hence, integrating by parts, and neglecting second order terms,

$$\begin{aligned} \int_0^T \left(\frac{\partial H}{\partial x} - \frac{\partial \bar{H}}{\partial x} \right) \tilde{x} dt &\approx -\varepsilon \int_0^T \Pi \left[\frac{\partial}{\partial t} \left(\frac{\partial H}{\partial x} (q, y, u_0(t, \cdot), t, \cdot) \right) \right] \left(\frac{t}{\varepsilon} \right) r_1 dt \\ &\quad + \varepsilon \int_0^T p_2^T \left(t, \frac{t}{\varepsilon} \right) \left[f \left(x, u, t, \frac{t}{\varepsilon} \right) - f \left(x - r_1, u - v, t, \frac{t}{\varepsilon} \right) \right] dt. \end{aligned}$$

But

$$\begin{aligned} &\varepsilon \int_0^T p_2^T \left(t, \frac{t}{\varepsilon} \right) \left[f \left(x, u, t, \frac{t}{\varepsilon} \right) - f \left(x - r_1, u - v, t, \frac{t}{\varepsilon} \right) \right] dt \\ &\quad \geq \varepsilon \int_0^T p_2^T \left(t, \frac{t}{\varepsilon} \right) \left[\frac{\partial f}{\partial x} r_1 + \frac{\partial f}{\partial u} v \right] dt - k\varepsilon \int_0^T \left| p_2^T \left(t, \frac{t}{\varepsilon} \right) \right| (|r_1|^2 + |v|^2) dt \\ &\quad \geq \varepsilon \int_0^T p_2^T \left(t, \frac{t}{\varepsilon} \right) \left[\frac{\partial f}{\partial x} r_1 + \frac{\partial f}{\partial u} v \right] dt - k\varepsilon (\varepsilon^2 + z_\varepsilon^2) \end{aligned}$$

by Lemma 1.2. On the other hand,

$$\begin{aligned} &\varepsilon \int_0^T \Pi \left[\frac{\partial}{\partial t} \left(\frac{\partial H}{\partial x} (q, y, u_0(t, \cdot), t, \cdot) \right) \right] \left(\frac{t}{\varepsilon} \right) r_1 dt \\ &\quad = \varepsilon \left\{ \int_0^T \Pi \left[\frac{\partial}{\partial t} \left(\frac{\partial H}{\partial x} (q, y, u_0(t, \cdot), t, \cdot) \right) \right] \left(\frac{t}{\varepsilon} \right) dt \right\} r_1(T) \\ &\quad \quad - \varepsilon \int_0^T \left\{ \int_0^t \Pi \left[\frac{\partial}{\partial t} \left(\frac{\partial H}{\partial x} (q, y, u_0(s, \cdot), s, \cdot) \right) \right] \left(\frac{s}{\varepsilon} \right) ds \right\} \frac{dr_1}{dt} dt \\ &\quad \geq -k\varepsilon^2 \left(\|r_1\|_\infty + \left\| \frac{dr_1}{dt} \right\|_{L^1} \right), \end{aligned}$$

by averaging estimations. But, from (9) and Lemmas 1.2 and 1.3, one sees that there exists $k > 0$ such that $\|dr_1/dt\|_{L^1} \leq k(1 + z_\varepsilon^2)$; this completes the proof.

We are now going to study the second order term with $H_t(\lambda, \mu)$ in Lemma 1.1.

LEMMA 1.5. *There exists $k > 0$ such that:*

$$\begin{aligned} &\int_0^1 dt \int_0^1 \lambda d\lambda \int_0^1 d\mu (\tilde{x}^T + \varepsilon x_2^T, \tilde{u}^T) H_t(\lambda, \mu) \begin{pmatrix} \tilde{x} + \varepsilon x_2 \\ v \end{pmatrix} \\ &\quad \geq 2\varepsilon \int_0^1 dt \int_0^1 \lambda d\lambda \int_0^1 d\mu (y_1 + v_1^T) H_t(\lambda, \mu) \begin{pmatrix} r + \varepsilon x_2 \\ v \end{pmatrix} - k\varepsilon^2 + \beta z_\varepsilon^2. \end{aligned}$$

Proof of Lemma 1.5. By substituting $Z_\varepsilon - [(\partial^2 H / \partial v^2)^{-1} \partial H / \partial v \partial x](r + \varepsilon x_2)$ to v in the integral and using Assumptions (H4) and (H5), one shows that the expression on the left-hand side is greater than $\int_0^1 dt \int_0^1 \lambda d\lambda \int_0^1 d\mu \beta |Z_\varepsilon(\lambda, \mu)(t)|^2$, that is, βz_ε^2 .

This will be the only positive term in z_ε^2 ; the others will be of the form $-k\varepsilon z_\varepsilon^2$.

Note that Lemma 1.4 already displays one part of the cost to be minimized in problem (TP^ε) . The other part will appear by replacing $H_t(\lambda, \mu)$ by $H_t(0, 0)$ in the estimates of Lemma 1.5.

LEMMA 1.6. *There exists $k \geq 0$ such that:*

$$\begin{aligned} 2\varepsilon \int_0^T dt \int_0^1 \lambda d\lambda \int_0^1 d\mu (y_1^T, v_1^T) H_t(\lambda, \mu) \begin{pmatrix} r + \varepsilon x_2 \\ v \end{pmatrix} \\ \geq \varepsilon \int_0^T (y_1^T, v_1^T) H_t(0, 0) \begin{pmatrix} r + \varepsilon x_2 \\ v \end{pmatrix} dt - k\varepsilon^2 - k\varepsilon z_\varepsilon^2. \end{aligned}$$

Proof of Lemma 1.6. Since the second derivatives of f and L are Lipschitz,

$$\|H_t(\lambda, \mu) - H_t(0, 0)\| \leq K\lambda\mu(|x - y|^2 + |u - u_0|^2)^{1/2} = K\lambda\mu(|r + \varepsilon x_2 + \varepsilon y_1|^2 + |v - \varepsilon v_1|^2)^{1/2}.$$

As x_2, y_1 and v_1 are bounded, there exists $k > 0$ such that:

$$\begin{aligned} \varepsilon \left| \int_0^T dt \int_0^1 \lambda d\lambda \int_0^1 d\mu (y_1 + v_1^T) [H_t(\lambda, \mu) - H_t(0, 0)] \begin{pmatrix} r + \varepsilon x_2 \\ v \end{pmatrix} \right| \\ \leq k\varepsilon(\varepsilon^2 + \|r\|_\infty^2 + \|v\|_L^2). \end{aligned}$$

We get the result from Lemma 2.

LEMMA 1.7. *There exists $k \geq 0$ such that:*

$$\begin{aligned} \int_0^T L(x, u, t) dt &\geq \int_0^T L(x, u^\varepsilon, t) dt + \varepsilon \int_0^T \overline{\frac{\partial H}{\partial x}} x_2 dt - \varepsilon q(0)^T x_2(0, 0) \\ &\quad + \varepsilon \int_0^T p_2^T \left(t, \frac{t}{\varepsilon} \right) \left[\frac{\partial f}{\partial x} r_1 + \frac{\partial f}{\partial u} v \right] dt \\ &\quad + \varepsilon \int_0^T (y_1^T, v_1^T) H_t(0, 0) \begin{pmatrix} r + \varepsilon x_2 \\ v \end{pmatrix} dt + z_\varepsilon^2(\beta - k\varepsilon) - k\varepsilon^2. \end{aligned}$$

Proof of Lemma 1.7. Combine the results of Lemmas 1.1, 1.4, 1.5 and 1.6.

We are going now to complete the estimate by using the problem (TP^ε) .

LEMMA 1.8. *There exists $k \geq 0$ such that:*

$$\varepsilon \int_0^T p_2^T \left(t, \frac{t}{\varepsilon} \right) \left[\frac{\partial f}{\partial x} r_1 + \frac{\partial f}{\partial v} v \right] dt + \varepsilon \int_0^T (y_1^T, v_1^T) H_t(0, 0) \begin{pmatrix} r + \varepsilon x_2 \\ v \end{pmatrix} dt \geq -k\varepsilon(\varepsilon^2 + z_\varepsilon^2).$$

Proof of Lemma 1.8. Transformations using the adjoint equations of (TP^ε) and the explicit value of v_1 as a feedback yield the following estimate for the expression on the left-hand side:

$$\varepsilon \int_0^T q_1^T \left[f \left(x, u, t, \frac{t}{\varepsilon} \right) - f \left(x - r_1, u - v, t, \frac{t}{\varepsilon} \right) - \frac{\partial f}{\partial x} r_1 - \frac{\partial f}{\partial u} v \right] dt$$

which is clearly of second order in $(\|r_1\|^2 + \|v\|_L^2)^{1/2}$; from Lemmas 1.2 and 1.3 we get the result.

From Lemmas 1.8 and 1.7 the estimation proposed in Lemma 1 is proved for ε sufficiently small to make βz_ε^2 dominant against $-k\varepsilon$. This completes the proof of Lemma 1.

LEMMA 2. *We are going to estimate the cost induced by U^ε .*

Let now $u = u^\varepsilon$, that is $\tilde{u} = 0$. Then:

$$\int_0^T L(x, u^\varepsilon, t) dt \approx \int_0^T L(y, u^\varepsilon, t) dt - \varepsilon q^T(0)x_2(0, 0) + \varepsilon \int_0^T \overline{\frac{\partial H}{\partial x}} x_2 dt.$$

Proof of Lemma 2. We can use Lemma 1.1 to get a first estimate:

$$\begin{aligned} \int_0^T L(x, u^\varepsilon, t) dt &\approx \int_0^T L(x, u^\varepsilon, t) dt - \varepsilon q^T(0)x_2(0, 0) + \varepsilon \int_0^T \overline{\frac{\partial H}{\partial x}} x_2 dt \\ &\quad + \int_0^T \left(\frac{\partial H}{\partial x} - \frac{\partial \bar{H}}{\partial x} \right) \tilde{x} dt \\ &\quad + \int_0^T dt \int_0^1 \lambda d\lambda \int_0^1 d\mu (\tilde{x}^T + \varepsilon x_2^T) \frac{\partial^2 H}{\partial x^2}(\lambda, \mu) (\tilde{x} + \varepsilon x_2) \end{aligned}$$

where $\partial^2 H / \partial x^2$ is computed at the same point as for $H_t(\lambda, \mu)$. Since \tilde{x} is of order one in ε , the last integral is of order two. Proceeding as in Lemma 1.4, we also show that the integral before that one is of order two. This completes the proof of Lemma 2. Theorem 1 follows from Lemmas 1 and 2.

COROLLARY 1. *If u is a "better" control than u^ε for the problem (P^ε) , then:*

$$\|x - y\|_\infty \leq k\varepsilon \left(1 + \frac{1}{\sqrt{\beta}} \right) \quad \text{and} \quad \|u - u_0\|_{L^2} \leq k\varepsilon \left(1 + \frac{1}{\sqrt{\beta}} \right).$$

Proof of Corollary 1. Denote by $J^\varepsilon(u)$ the cost in problem (P^ε) . For $\varepsilon \in]0, 1]$, one has $J^\varepsilon(u) \geq J^\varepsilon(u^\varepsilon) - k\varepsilon^2 + (\beta - k\varepsilon)z_\varepsilon^2$, with $k \geq 0$. If u is better than u^ε , then $(\beta - k\varepsilon)z_\varepsilon^2 \leq k\varepsilon^2$. Take $\varepsilon < \beta/2k$; then $z_\varepsilon^2 \leq 2k/\beta\varepsilon^2$. The result follows from Lemma 1.2.

2. The nonperiodic case.

2.1. An ergodic theorem on O.D.E. Let

$$f: \begin{cases} \mathbf{R}^n \times [0, T] \times \mathbf{R}_+ \rightarrow \mathbf{R}^n \\ (x, t, \theta) \rightarrow f(x, t, \theta) \end{cases}$$

meeting the following assumptions:

(H7) f is Lipschitz in (x, t) with Lipschitz constant Λ , and integrable in θ .

(H8) f has an average \bar{f} in the sense that, for any bound B , one has:

$$\sup_{\substack{|x| \leq B \\ t \in [0, T] \\ \hat{t} \geq 0}} \left| \frac{1}{\tau} \int_{\hat{t}}^{\hat{t}+\tau} f(x, t, \theta) d\theta - \bar{f}(x, t) \right| \xrightarrow{\tau \rightarrow +\infty} 0.$$

As before, (H7) ensures that one has a true separation of time scales.

Let x^ε and y be defined by:

$$(10) \quad \frac{dx^\varepsilon}{dt} = f\left(x^\varepsilon, t, \frac{t}{\varepsilon}\right), \quad x^\varepsilon(0) = x_0, \quad t \in [0, T],$$

$$(11) \quad \frac{dy}{dt} = \bar{f}(y, t), \quad y(0) = x_0, \quad t \in [0, T].$$

Then

$$\sup_{t \in [0, T]} |x^\varepsilon(t) - y(t)| \xrightarrow{\varepsilon \rightarrow 0} 0.$$

Proof. At a fast time scale, the slow time t can be considered as constant, as well as $x^\varepsilon(t)$ and $y(t)$. Hence the dynamics f can be approximated by \bar{f} ; integration yields the result.

More precisely, let $D \in \mathbb{N}^*$ and, for $t \in [0, T]$, let $t_k = kt/D$. Then:

$$\begin{aligned} |x^\varepsilon(t) - y(t)| &\leq \Lambda \int_0^t |x^\varepsilon(s) - y(s)| ds + \sum_{k=0}^{D-1} \int_{t_k}^{t_{k+1}} \left| f\left(y(s), s, \frac{s}{\varepsilon}\right) - f\left(y(t_k), t_k, \frac{s}{\varepsilon}\right) \right| ds \\ &\quad + \sum_{k=0}^{D-1} \int_{t_k}^{t_{k+1}} |\bar{f}(y(s), s) - \bar{f}(y(t_k), t_k)| ds \\ &\quad + \sum_{k=0}^{D-1} \int_{t_k}^{t_{k+1}} \left| f\left(y(t_k), t_k, \frac{s}{\varepsilon}\right) - \bar{f}(y(t_k), t_k) \right| ds \\ &\leq \Lambda \int_0^t |x^\varepsilon(s) - y(s)| ds + \Lambda_1 \frac{t^2}{D} + t \sup_{\substack{|x| \leq \|y\|_\infty \\ t \in [0, T] \\ \hat{t} \geq 0}} \left| \frac{D\varepsilon}{t} \int_{\hat{t}}^{\hat{t}+\tau} f(x, t, \theta) d\theta \right. \\ &\quad \left. - \bar{f}(x, t) \right| \xrightarrow{\tau \rightarrow +\infty} 0. \end{aligned}$$

Choose $D = D(\varepsilon)$ such that $D(\varepsilon) \rightarrow_{\varepsilon \rightarrow 0} \infty$ and $\varepsilon D(\varepsilon) \rightarrow_{\varepsilon \rightarrow 0} 0$; use of the Gronwall lemma yields the result.

2.2. The averaged problem. Section 2.1 can be viewed as a generalization of averaging techniques to the nonperiodic case. We are now going to use it to define the averaged problem in the nonperiodic case.

Let f and L be as in § 1.1, except that now they need not be periodic in θ , and define problem (P^ε) accordingly. The important point in the definition of the averaged problem is that of the set W^{ad} of admissible controls. W^{ad} will be the set of functions u from $[0, T] \times \mathbb{R}_+$ to \mathbb{R}^p , with values in U^{ad} for almost every (t, θ) , and such that $f(x, u(t, \theta), t, \theta)$ and $L(x, u(t, \theta), t, \theta)$ have an average in θ for every (x, t) .

Note that W^{ad} may be empty. However, we will show that, if averaging can reasonably be expected to be used (i.e., the minimized Hamiltonian has an average), then W^{ad} is nonempty (see § 2.3).

For u in W^{ad} we define the averaged problem:

$$\left. \begin{aligned} \frac{dy}{dt} &= \bar{f}(y, u(t, \cdot), t, \cdot), \quad y(0) = x_0, \quad t \in [0, T], \\ \text{Minimize } \int_0^T \bar{L}(y, u(t, \cdot), t, \cdot) dt, \end{aligned} \right\} (\bar{P}).$$

If f and L are periodic, we find the same definition as in § 1.1.

2.3. The Hamiltonian of the averaged problem. In this section, we will omit the mention x and t . All assumptions made will be supposed to hold for every x and t .

Define the pseudo-Hamiltonian $h(p, u, \theta)$ by $h(p, u, \theta) = p^T f(u, \theta) + L(u, \theta)$ and let $H(p, \theta) = \min_{u \in U^{ad}} h(p, u, \theta)$ when it exists.

We are going to show that if H has an average for any p , then its average is the minimum of the Hamiltonian of the averaged problem. We will use the following assumptions:

- (H9) For any bounded part B of \mathbf{R}^p , f and L are bounded and uniformly continuous on $B \times \mathbf{R}_+$ (i.e., f and L are in $\text{BUC}(B, \mathbf{R}_+)$).
- (H10) For any (p, θ) , $h(p, u, \theta)$ has a minimum on U^{ad} . Moreover, for any bounded domain B in \mathbf{R}^p there exists a compact set K in U^{ad} such that the minimum can be reached in K for any (p, θ) in $B \times \mathbf{R}_+$.
- (H11) H has an average \bar{H} , i.e.,

$$\frac{1}{\tau} \int_0^\tau H(p, \theta) d\theta \xrightarrow{\tau \rightarrow +\infty} \bar{H}(p) \quad \text{for all } p \text{ in } \mathbf{R}^n.$$

THEOREM 2. *Let f and L meeting (H9), (H10) and (H11). Then W^{ad} is nonempty and $\bar{H}(p) = \text{Min}_{v \in W^{ad}} p^T \bar{f}(v) + \bar{L}(v)$.*

Proof. Denote $p^T \bar{f}(v) + \bar{L}(v)$ by $\bar{h}(p, v)$ for v in W^{ad} and let $E = \{p \in \mathbf{R}^n \mid \exists v \in W^{ad}, \bar{H}(p) = \bar{h}(p, v)\}$. As obviously $\bar{h}(p, v) \geq \bar{H}(p)$ for any v in W^{ad} , Theorem 2 is equivalent to $E = \mathbf{R}^n$. We are going to show that E is closed and that $\mathbf{R}^n - E$ is of null measure (Lebesgue).

(i) E is closed.

If $E = \emptyset$, this is true. If $E \neq \emptyset$, let p_n be a sequence in E , converging in \mathbf{R}^n , with $\bar{H}(p_n) = \bar{h}(p_n, v_n)$. Thanks to (H9) and (H10), $\bar{f}(v_n)$ and $\bar{L}(v_n)$ are bounded; let \bar{f} and \bar{L} be two cluster points, and w_n a subsequence such that $\bar{f}(w_n) \rightarrow_{n \rightarrow \infty} \bar{f}$ and $\bar{L}(w_n) \rightarrow_{n \rightarrow \infty} \bar{L}$. We are going to exhibit a control v such that $p^T \bar{f} + \bar{L} = p^T \bar{f}(v) + \bar{L}(v)$.

Let τ_n be an increasing sequence in \mathbf{R}_+ such that $\tau_n \geq n!$ and such that, for $\tau \geq \tau_n$:

$$\left| \frac{1}{\tau} \int_0^\tau f(w_n(\theta), \theta) d\theta - \bar{f}(w_n) \right| < \frac{1}{n}.$$

τ_n exists since w_n is in W^{ad} . Define v by $v(\theta) = w_n(\theta)$ for θ in $[\tau_n, \tau_{n+1}[$; it is then easy to check that v is in W^{ad} and that $p^T \bar{f} + \bar{L} = \bar{h}(p, v)$.

(ii) $\mathbf{R}^n - E$ is of null measure.

\bar{H} is locally Lipschitz, thanks to Assumptions (H9) and (H10), and thus, if F denotes the set of differentiability points of \bar{H} , $\mathbf{R}^n - F$ is of null measure. We are going to show that $F \subset E$.

Notice that, thanks to (H9) and (H10), there exists a measurable function u from $\mathbf{R}^n \times \mathbf{R}_+$ to U^{ad} such that $H(p, \theta) = h(p, u(p, \theta), \theta)$. Let ρ be a small positive number, p in F and q a direction in \mathbf{R}^n . Then:

$$\frac{H(p + \rho q) - H(p, \theta)}{\rho} \leq q^T f(p, u(p, \theta), \theta) \leq \frac{H(p, \theta) - H(p - \rho q, \theta)}{\rho}.$$

Let l be a cluster point of $1/\tau \int_0^\tau q^T f(p, u(p, \theta), \theta) d\theta$ as $\tau \rightarrow +\infty$; l exists thanks to (H9) and (H10). Then let $\tau \rightarrow \infty$ first, then $\rho \rightarrow 0$ in the above inequalities. We conclude that $l = \partial \bar{H} / \partial p$ in the direction q .

As this is true for any cluster point and any direction q , we have

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau f(p, u(p, \theta), \theta) d\theta = \frac{\partial \bar{H}^T}{\partial p}.$$

Since f and H have an average, L has an average and $u(p, \theta)$ is in W^{ad} , with $\bar{H}(p) = p^T \bar{f}(u) + \bar{L}(u)$.

Remark 5. The concavity of H in p is essential. Let $H(p, \theta) = \sin(p, \theta)$. $\bar{H} = 0$, yet $\partial H / \partial p$ has no average.

2.4. A limit theorem on the Bellman function. Assumption (H11) has given sense to the averaged problem by ensuring that W^{ad} is nonempty. It also yields a limit theorem on the Bellman function. We will consider the latter as the unique viscosity solution of the Hamilton–Jacobi–Bellman equation (see [7]).

Assumptions. Let

$$H: \begin{cases} \mathbf{R}^n \times \mathbf{R}^n \times [0, T] \times \mathbf{R}_+ \rightarrow \mathbf{R}, \\ (p, x, t, \theta) \rightarrow H(p, x, t, \theta). \end{cases}$$

H may represent, for instance, the minimized Hamiltonian of § 2.3. The assumptions are the following:

$$(H12) \quad H \in \text{BUC}(B \times \mathbf{R}^n \times [0, T] \times \mathbf{R}_+) \text{ for any bounded part } B \text{ of } \mathbf{R}^n, \text{ and } |H(p, x, t, \theta) - H(p, y, t, \theta)| \leq C(1 + |p|)(|x - y|).$$

$$(H13) \quad H \text{ has an average } \bar{H} \text{ such that, for any } p, x, t,$$

$$\sup_{\hat{t} \geq 0} \left| \frac{1}{\tau} \int_{\hat{t}}^{\hat{t} + \tau} H(p, x, t, \theta) d\theta - \bar{H}(p, x, t) \right| \xrightarrow{\tau \rightarrow +\infty} 0.$$

THEOREM 3. *Let V be the viscosity solution of:*

$$(12) \quad \frac{\partial V}{\partial t} + \bar{H}\left(\frac{\partial V}{\partial x}, x, t\right) = 0, \quad V(x, T) \equiv 0, \quad x \in \mathbf{R}^n, \quad t \in [0, T].$$

Let V^ε be the viscosity solution of:

$$(13) \quad \frac{\partial V^\varepsilon}{\partial t} + H\left(\frac{\partial V^\varepsilon}{\partial x}, x, t, \frac{t}{\varepsilon}\right) = 0, \quad V^\varepsilon(x, T) \equiv 0, \quad x \in \mathbf{R}^n, \quad t \in [0, T].$$

Then V^ε converges to V as ε tends to zero, uniformly on any compact subset of $\mathbf{R}^n \times [0, T]$.

Proof. We will make use of the following notation:

$$L^p_{\text{unif}}(\mathbf{R}^n \times [0, T]) = \left\{ \phi \in L^p_{\text{loc}}(\mathbf{R}^n \times [0, T], \mathbf{R}), \sup_{y \in \mathbf{R}^n} \int_{\substack{|x-y| < 1 \\ t \in [0, T]}} |\phi(x, t)|^p dx dt < \infty \right\}.$$

$W^{2,1,p}_{\text{unif}}(\mathbf{R}^n \times [0, T])$ is the set of functions ϕ in $L^p_{\text{unif}}(\mathbf{R}^n \times [0, T])$ such that $\partial \phi / \partial t$, $\partial \phi / \partial x$ and $\partial^2 \phi / \partial x^2$ exist and are in L^p_{unif} . $W^{2,1,p}_{\text{unif}}$ is the analogue of $W^{2,1,p}$ except that L^p is replaced by L^p_{unif} . We can use the norm on $W^{2,1,p}_{\text{unif}}$ defined by the sup of $W^{2,1,p}$ norms over all $\bar{B}(y, 1) \times [0, T]$, where $\bar{B}(y, 1)$ is the closed ball of center y and radius 1.

We will denote, for $\alpha > 0$, by V^ε_α the unique solution in $\cap_{p \geq 0} W^{2,1,p}_{\text{unif}}$ of

$$(14) \quad \frac{\partial V^\varepsilon_\alpha}{\partial t} + \alpha \Delta V^\varepsilon_\alpha + H\left(\frac{\partial V^\varepsilon_\alpha}{\partial x}, x, t, \frac{t}{\varepsilon}\right) = 0, \quad V^\varepsilon_\alpha(x, T) \equiv 0$$

and V_α the analogue for

$$(15) \quad \frac{\partial V_\alpha}{\partial t} + \alpha \Delta V_\alpha + \bar{H}\left(\frac{\partial V_\alpha}{\partial x}, x, t\right) = 0, \quad V_\alpha(x, T) \equiv 0.$$

It is well known [7] that

$$\|V^\varepsilon_\alpha - V^\varepsilon\|_\infty \leq k\sqrt{\alpha} \quad \text{and} \quad \|V_\alpha - V\|_\infty \leq k_2\sqrt{\alpha}.$$

Moreover, k does not depend on the behaviour of H in time, that is, in particular k does not depend on ε . Hence, the theorem will be proved if, α being fixed, V_α^ε converges to V_α for ε in $]0, 1]$. Estimates on the derivatives of V_α^ε ensure that the V_α^ε are bounded in (and thus form a weakly relatively compact subset of) $W_{\text{unif}}^{2,1,p}$ for any $p > 0$. Thus, the $(V_\alpha^\varepsilon, \partial V_\alpha^\varepsilon / \partial x)$ are relatively compact in $C^0(K \times [0, T], \mathbf{R})$, K being any compact subset of \mathbf{R}^n . Let $(\bar{V}_\alpha, \partial \bar{V}_\alpha / \partial x)$ a cluster point and $(V_{\alpha_n}^\varepsilon, \partial V_{\alpha_n}^\varepsilon / \partial x)$ a sequence converging to $(\bar{V}_\alpha, \partial \bar{V}_\alpha / \partial x)$ uniformly on any compact, with $V_{\alpha_n}^\varepsilon \rightarrow_{n \rightarrow \infty} \bar{V}_\alpha$ in $W_{\text{unif}}^{2,1,p}$ weakly.

Proving the theorem thus amounts to showing that \bar{V}_α is a solution of (15) in the weak sense. Let ϕ a function in $C^\infty(\mathbf{R}^n \times [0, T])$ with a compact domain.

$$\begin{aligned} \int_{t_k}^{t_{k+1}} dt \int_{\mathbf{R}^n} dx & \left[\frac{\partial \bar{V}_\alpha}{\partial t} + \alpha \Delta \bar{V}_\alpha + \bar{H} \left(\frac{\partial \bar{V}_\alpha}{\partial x}, x, t \right) \right] \phi(x, t) \\ &= \int_{t_k}^{t_{k+1}} dt \int_{\mathbf{R}^n} dx \left[\frac{\partial \bar{V}_\alpha}{\partial t} - \frac{\partial V_\alpha^\varepsilon}{\partial t} + \alpha \Delta \bar{V}_\alpha - \alpha \Delta V_\alpha^\varepsilon \right] \phi(x, t) \\ & \quad + \int_{t_k}^{t_{k+1}} dt \int_{\mathbf{R}^n} dx \left[H \left(\frac{\partial \bar{V}_\alpha}{\partial x}, x, t, \frac{t}{\varepsilon} \right) - H \left(\frac{\partial V_\alpha^\varepsilon}{\partial x}, x, t, \frac{t}{\varepsilon} \right) \right] \phi(x, t) \\ & \quad + \int_{t_k}^{t_{k+1}} dt \int_{\mathbf{R}^n} dx \left[\bar{H} \left(\frac{\partial \bar{V}_\alpha}{\partial x}, x, t \right) - H \left(\frac{\partial \bar{V}_\alpha}{\partial x}, x, t, \frac{t}{\varepsilon} \right) \right] \phi(x, t). \end{aligned}$$

As ε tends to zero, the first expression has limit zero by weak convergence in $W_{\text{unif}}^{2,1,2}$. The second one tends to zero thanks to the Lebesgue dominated convergence theorem. A discretization scheme similar to that used in § 2.1 ensures that the third one has limit zero, thanks to Assumption (H13).

Remark 6. It can be proved [4] that, if H is locally Lipschitz in p and if the convergence in Assumption (H13) is uniform for x in \mathbf{R}^n , then the convergence of V^ε to V is uniform on $\mathbf{R}^n \times [0, T]$.

One may wonder when (H12) is true with $H = p^T f + L$. It is true if, for instance, f and L are BUC and Lipschitz in x . However, this can be extended to the case where f is uniformly continuous, Lipschitz in x with $|f| \leq k(1 + |x| + |u|)$; and L is uniformly continuous, locally Lipschitz in x , $|\partial L / \partial x| \leq k(1 + |x| + |u|)$ and $|L| \leq k(1 + |x|^2 + |u|^2)$, $L \geq k_1(-1 + |u|^2)$.

This includes the linear quadratic case; actually, we might call this the “sublinear quadratic case.” With a suitable truncation of f and L on the phase domain, we may keep V^ε and V unchanged on a portion of $\mathbf{R}^n \times [0, T]$, while retrieving Assumption (H12). The convergence is still uniform on any compact.

3. Perspectives. We have proven here, at least in the periodic case, that averaging can be used as an efficient tool in deterministic optimal control. It is efficient for two reasons.

First, the averaged problem (\bar{P}) is easier to solve numerically than the original problem (P^ε) , since a “fast” time grid is no longer needed in the simulation part. Gains should also be expected on the state space grid, since it is often related to the former one; it is an important point if one thinks, for instance, of dynamic programming.

Second, and thanks to Theorem 1, the solution of (\bar{P}) is known to be near optimal for (P^ε) ; hence, we do not lose much by solving the averaged problem instead of the original one.

We have shown that use of the averaged problem can also be expected to be efficient in the nonperiodic case, as the optimal cost of (P^ε) is close to that of (\bar{P}) when ε is small (Theorem 3).

However, practical problems are seldom under the form (P^ε) , be it in the periodic or nonperiodic case. Most of the time, there is, for instance, no explicit separation of the time scales under the form (t, θ) , with θ ranging from 0 to $+\infty$; in particular, periodicity or averaging assumptions cannot be checked directly. Nevertheless, the results presented here provide an important theoretical background for developments of both theoretical and practical interest.

From the theoretical point of view, it is reasonable to expect problem (TP^ε) in Lemma 1 to provide further expansions of the cost (J^ε) , as similar methods have already been used with success in the case of regular and singular perturbations [2]. A complete expansion of the Bellman function in the linear quadratic periodic case has already been obtained [4]. In particular, the terms of order higher than two are in the form $V(x, t, t/\varepsilon, (T - t/\varepsilon))$, with periodicity in both the forward and backward fast times. This is probably related to the existence of the terms x_2 and p_2 in the expansion of the primal and dual trajectories. The same phenomenon exists in singular perturbations with boundary layers instead of phase terms.

We have seen that x_2 and p_2 are defined through the operator Π . In fact, Π appears in any expansion of an integral with an integrand periodic in fast time. A generalization of Π would be welcome if we hope to find some results equivalent to Theorem 1 in the nonperiodic case.

At last, links should be developed with singular perturbations. From the practical point of view, we have seen that the assumptions in Theorems 1 and 3 cannot be checked directly. It should be noticed (especially in the nonperiodic case) that the question is not so much that the assumptions might not hold, as it is rather to immerse the optimization problem in the "right" family of problems (P^ε) . Moreover, the ideas are sufficiently simple and general to be used in heuristics. One can think, for instance, of separating the time scales through the use of "moving averages." Heuristics can also be developed to generalize the operator Π to the nonperiodic case and use it to improve performances. We are going to experiment numerically on these ideas.

Conversely, averaging has been often used empirically by engineers in practical problems. The results and notions presented here may provide them some guidelines in further applications.

We have discussed *how* averaging could be used practically. We shall discuss now *when* it could be used. Heuristically, averaging can be expected to yield good results and performances when the dynamics of a system depend on a fast "erratic" exogenous phenomenon. A good example is given by weather disturbances.

However, these phenomena are often modeled by stochastic processes. As we mentioned before, both approaches are similar in the sense that they make use of averages; their theoretical background is, however, quite different. Moreover, averaging has also been used in stochastic control ([5], for instance). In all cases, the original data consists often of a finite number of physical measures, in no way probabilistic or two-time scaled in nature. Thus, neither approach is justified a priori. Therefore, it should prove quite interesting to experiment all methods on various sets of data. We plan to conduct such experiments.

REFERENCES

- [1] V. ARNOLD, *Chapitres supplémentaires de la théorie des équations différentielles ordinaires*, French translation from the Russian, Mir, Moscow, 1980.
- [2] A. BENSOUSSAN, *Singular Perturbations in Systems and Control*, Mark Ardem, ed., Springer-Verlag, Berlin-Heidelberg-New York, 1983, pp. 169-185.

- [3] A. BENSOUSSAN, J. L. LIONS AND G. PAPANICOLAOU, *Asymptotic Analysis for Periodic Structures*, North-Holland, Amsterdam, 1978.
- [4] F. CHAPLAIS, *Averaging et contrôle optimal déterministe*, Thèse de Docteur-Ingénieur, Ecole nationale Supérieure des Mines de Paris, Paris, 1984.
- [5] F. DELBECQUE AND J. P. QUADRAT, *Contribution of Stochastic Control Singular Perturbation Averaging and Team Theory to an Example of Large-Scale Systems: Management of Hydropower Production*, IEEE Trans. Automat. Control, AC-23, April 1978, pp. 209–221.
- [6] P. V. KOKOTOVIC, R. E. O'MALLEY, JR. AND P. SANNUTI, *Singular perturbations and order reduction in control theory: An overview*, Automatica, 12 (1976), pp. 123–132.
- [7] P. L. LIONS, *Generalized Solutions of Hamilton–Jacobi Equations*, Pitman, Boston, 1982.